

Resumen Ejecutivo2 estandarización y agrupamiento de datos

Importación de datos

Explotamos la pandas biblioteca para importar el conjunto de datos y lo transformamos en un marco de datos a través de la read_csv()función

```
importar pandas como pd df = pd.read_csv('cupcake.csv') df.head(5)
```

Agrupación por distancia

En este caso definimos los bordes de cada contenedor. En Python, pandas la agrupación por distancia se logra mediante la cut ()función.

```
valor_min = df['Cupcake'].min() valor_max = df['Cupcake']. max() print(valor_min)  
print(valor_max)
```

Agrupamos los valores relacionados con la columna Cupcakeen tres grupos: pequeño, mediano y grande.

La agrupación de datos es un tipo de preprocesamiento de datos, un mecanismo que incluye también el manejo de valores faltantes, normalización y estandarización.

convertir numérico a categórico incluye agrupamiento por distancia y agrupamiento por frecuencia reducir valores numéricos incluye la cuantificación (o muestreo)

Agrupación por distancia

Con medio de este código lo podemos correr.

```
import matplotlib.pyplot as plt  
  
plt.hist(df['bins'], bins=3)
```